

\*\*\*\*\*  
M P S E R E H  
(TM)  
\*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSPrch\_PP protein - protein database search, using Smith-Waterman algorithm  
Rt: Sun Jun 4 15:31:32 2000; MasPar time 27.68 Seconds  
Tabular output not generated. 698,804 Million cell updates/sec

Title: >US-09-471-749-5  
Description: (1-410) from US09471749.pep  
Perfect Score: 3005  
Sequence: 1 MLDQVSEYLGVTGTSFRKRYP.....QAPPPPKYGRGKNSKEG 410

Scoring table: PAM 150  
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: plf62  
1:plf1 2:plf2 3:plf3 4:plf4

Statistics: Mean 48.931; Variance 106.509; scale 0.459

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

R	C	Score	Query Match	Length	ID	Description	Pred. No.
1	436	14.5	371	2	A55302	probable transcriptio	2.00e-53
2	341	11.3	397	2	S26731	neuro-D4 protein - ra	9.49e-37
3	257	8.6	5262	2	T03454	ALR protein - human	1.33e-22
4	253	8.4	4957	2	T03455	ALR protein - human	5.95e-22
5	184	6.1	811	2	T08738	hypothetical protein	4.05e-11
6	176	5.9	1722	2	T78879	retinodiasoma bindin	6.25e-10
7	162	5.4	564	2	I48776	Smcy - mouse (fingern	6.75e-08
8	162	5.4	1033	2	I48775	Smcx protein (escapes	6.75e-08
9	162	5.4	1560	2	I54361	escapes X-chromosome	6.75e-08
10	160	5.3	728	2	SS7142	hypothetical protein	1.30e-07
11	154	5.1	449	2	T13495	hypothetical protein	9.15e-07
12	149	5.0	1250	2	T00454	hypothetical protein	4.54e-06
13	148	4.9	1214	2	JC2069	zinc-finger protein,	6.24e-06
14	137	4.6	3866	2	I53035	zinc-finger protein,	1.93e-04
15	137	4.6	3866	2	B48205	trithorax homolog Hrx	1.93e-04
16	137	4.6	3869	2	A48205	All-1 protein +GRE fo	1.93e-04
17	138	4.6	3910	2	A44265	trithorax homolog Hrx	1.42e-04
18	138	4.6	3968	2	A44265	trithorax homolog Hrx	1.42e-04
19	134	4.5	288	2	I38485	BCR11Q23 - human (f	4.81e-04
20	133	4.4	653	2	T01274	hypothetical protein	1.19e-03
21	131	4.4	748	2	SS4505	hypothetical protein	1.19e-03
22	133	4.4	825	2	T02518	hypothetical protein	6.50e-04
23	130	4.3	530	2	I38558	Mt-2 autoantigen 240	1.60e-03

24	123	4.1	401	2	S28653	hypothetical protein	1.25e-02
25	123	4.1	1051	2	S55259	trif protein - mouse	1.25e-02
26	121	4.0	283	2	T11644	hypothetical protein	2.23e-02
27	120	4.0	357	2	JC4090	FK506-binding 39k pro	2.96e-02
28	116	3.9	417	2	S35784	glycoprotein gd - bov	9.17e-02
29	117	3.9	429	2	A40452	keratin 21, type I, c	6.93e-02
30	114	3.8	330	2	S58255	hypothetical protein	1.60e-01
31	112	3.7	373	2	E64486	ATP-dependent 26S pro	4.77e-01
32	110	3.7	933	2	S41539	fibrinogen-binding pr	8.16e-01
33	108	3.6	207	2	T08109	oleosin-like protein	6.25e-01
34	109	3.6	361	2	S61312	cap protein 1, macro	8.16e-01
35	108	3.6	417	1	VGB818	glycoprotein D precu	8.16e-01
36	107	3.6	852	2	T06310	hypothetical protein	1.07e+00
37	109	3.6	1021	2	S26985	probable DNA-directed	1.07e+00
38	107	3.6	1618	2	S21424	nestin - human	8.16e-01
39	108	3.6	2588	2	T14342	NSD1 protein - mouse	1.39e+00
40	106	3.5	570	2	S52765	secd protein - Strept	1.39e+00
41	106	3.5	596	2	I38228	Shb protein - human	1.39e+00
42	106	3.5	639	2	S39242	Hmg protein - Madagas	1.39e+00
43	105	3.5	651	2	S18874	nucleolin - African c	1.80e+00
44	105	3.5	1191	2	T13850	gene u-shaped protein	1.80e+00
45	105	3.5	1257	2	T01020	hypothetical protein	1.80e+00

## ALIGNMENTS

RESULT 1  
ENTRY A55302 #type complete  
TITLE probable transcriptio factor regulam - mouse  
ORGANISM Mus musculus #common\_name house mouse  
DATE 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 10-Sep-1997

ACCESSIONS A55302  
REFERENCE Gablig, T.G.; Mantel, P.L.; Rosli, R.; Crean, C.D.  
#authors J. Biol. Chem. (1994), 269:29515-29519  
#journal Regulam: a novel zinc finger gene essential for apoptosis in  
#title myeloid cells.

#cross-references MIM:95050794  
#accession A55302  
#status preliminary  
#molecule\_type mRNA  
##residues 1-371 #label GAP

SUMMARY  
#cross-references GB:010435; NID:9606660; PID:9606661  
#length 371 #molecular-weight 42062 #checksum 5627

Query Match 14.5%; Score 436; DB 2; Length 371;  
Best Local Similarity 40.3%; Pred. No. 2.00e-53;  
Matches 56; Conservative 31; Mismatches 48; Indels 4; Gaps 3;

DB	216	EEGEDKEDSRPTPVSRSEQSKKPGDGLAPNNYCDGKDGKINKKGOPELVSC	275
QY	256	QKSKKAAK-PRKDGKRYVLSKVPKRYIPNAIGICL-KKESKKAASLINC	313
DB	276	SGCGNSGHESCLOFTPVMAAVKTYRMWCIECKCNLCGTSENDDQLFCDDCDRGYHMY	335
QY	314	SGCGNSGHESCIDNMEIYSMIKTYPMQMECKTCICQAPHEEDMRCDCDGYHNF	373

DB 336 CLTPMSPEPDSNSCHLC 354  
QY 374 CV-GIGALPSGRWICDC 390

## RESULT 2

ENTRY S26731 #type complete  
TITLE neuro-D4 protein - rat  
ORGANISM Rattus norvegicus #common\_name Norway rat  
DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Mar-1999

ACCESSIONS S26731  
REFERENCE Buchman, V.L.; Ninkina, N.N.; Bogdanov, Y.D.; Bortvin, A.L.;  
#authors Akopian, H.N.; Kiselev, S.L.; Krylova, O.Y.; Anokhin, K.V.;



```

Db      281  PCKMTMOLRRN-HSST-QFMNYSYCRCSNREDEVKFLKDCGSDNTHIFCLPPLSEV 338
QY      323  SCLDMTMELVSMIKITPYMOCMECKTCITICGPHHEEMFCDMCDRGYHTFCV--GLGAI 380
Db      339  PKGWRCPRC 348
QY      381  PSGRMICDC 390

RESULT      8
ENTRY      148775      #type fragment
TITLE      Smcx protein (escapes X-chromosome inactivation) - mouse
              (fragment)
ORGANISM    European house mouse
DATE        12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
              06-Dec-1996
ACCESSIONS  148775; S44139
REFERENCE    138235
#authors    Aguilnik, A.I.; Mitchell, M.J.; Mattei, M.G.; Borsani, G.;
              Ayner, P.A.; Lerner, J.L.; Bishop, C.E.
#journal    Hum. Mol. Genet. (1994) 3:879-884
#title      A novel X gene with a widely transcribed Y-linked homologue
              escapes X-inactivation in mouse and human.
#cross-references MUID:95038739
#accession  148775
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-1033 #label RES
#cross-references EMBL:Z29651; NID:g603861; PID:g603862
GENETICS
#gene       Smcx
SUMMARY     #length 1033 #checksum 8222

Query Match      5.4%; Score 162; DB 2; Length 1033;
Best Local Similarity 34.3%; Pred. No. 6,75e-08;
Matches 24; Conservative 13; Mismatches 29; Indels 4; Gaps 3

Db      269  PCKMTMML-RRNHNA-QFISYVCRCNCRDEDDKLLKDCGDNTHIFCLPPLPEI 326
QY      323  SCLDMTMELVSMIKITPYMOCMECKTCITICGPHHEEMFCDMCDRGYHTFCV--GLGAI 380
Db      327  PKGWRCPRC 336
QY      381  PSGRMICDC 390

RESULT      9
ENTRY      154361      #type complete
TITLE      escapes' X-chromosome inactivation - human
              #formal_name Homo sapiens #common_name man
ORGANISM    02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE        28-Feb-1997
ACCESSIONS  154361
REFERENCE    154361
#authors    Wu, J.; Ellison, J.; Salido, E.; Yen, P.; Mohandas, T.;
              Shapiro, L.J.
#journal    Hum. Mol. Genet. (1994) 3:153-160
#title      Isolation and characterization of XE169, a novel human gene
              that escapes X-inactivation.
#cross-references MUID:94214434
#accession  154361
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-1560 #label RES
#cross-references GB:I25270; NID:g457136; PID:g457137
GENETICS
#gene       XE169
SUMMARY     #length 1560 #molecular_weight 175804 #checksum 4506

Query Match      5.4%; Score 162; DB 2; Length 1560;
Best Local Similarity 34.3%; Pred. No. 6,75e-08;

```

[illegible]

```

RESULT 12
ENTRY 12
TITLE 12
ORGANISM 12
DATE 12
ACCESSIONS 12
REFERENCE 12
#authors 12
#submission 12
#accession 12
#status 12
#molecule_type 12
#residues 12
#cross-references 12
GENETICS 12
#gene 12
#map_position 12
#introns 12
SUMMARY 12
Query Match 12
Best Local Similarity 12
Matches 12
DB 12
QY 12
OY 12
RESULT 13
ENTRY 13
TITLE 13
ALTERNATE_NAMES 13
ORGANISM 13
DATE 13
ACCESSIONS 13
REFERENCE 13
#authors 13
#journal 13
#title 13
#cross-references 13
#accession 13
#molecule_type 13
#residues 13
#cross-references 13
COMMENT 13
CLASSIFICATION 13
KEYWORDS 13
FEATURE 13
653-708
23,28,41,44
#domain bromodomain homology #label BRO\
#binding_site zinc (Cys, Cys, His, His) #status
predicted\

```

```

120,205,462      #binding site phosphate (Ser) (covalent) (by casein
276,279,293,296  kinase II) #status predicted\
301,304,317,320  #binding-site zinc (Cys) #status predicted\
330,333,350,353  #binding-site zinc (His, Cys, Cys) #status
386,389,401,405  #binding-site zinc (Cys, Cys, His, Cys) #status
410,413,444,447  #binding-site zinc (His, Cys, Cys, His) #status
SUMMARY          #length 1214 #molecular-weight 137484 #checksum 1329

Query Match      4.98; Score 148; DB 2; Length 1214;
Best Local Similarity 38.0%; Pred. No. 6.24e-06;
Matches 19; Conservative 14; Mismatches 15; Indels 2; Gaps 1;

C 276 CCTCNDGCONSNVILFCMDONLAVHQCAGVPIPEGOMTCRCRLGSPS 325
348 CIIIC--GDPHHEEMFCMDKDRGYHFCVGLAIPSGRWICDCCQRAP 395

RESULT 14
ENTRY      I53035      #type fragment
TITLE      trithorax homolog HTX - human (fragment)
ORGANISM   Homo sapiens #common_name man
DATE       01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change
29-Aug-1997

ACCESSIONS I53035
REFERENCE   I53035
#authors    Mbangkollo, D.; Burnett, R.; McCabe, N.; Thirman, M.; Gill,
#journal    H.; Yu, H.; Rowley, J.D.; Diaz, M.O.
#title      DNA Cell Biol. (1995) 14:475-483
#cross-references GB:578570; NID:91042096; PID:91042097
#cross-references MIMD:95322025
#accession  I53035
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-366 #label RES
#cross-references GB:578570; NID:91042096; PID:91042097
GENETICS
#gene       GDB:MLL
#cross-references GDB:128819; OMIM:159555
#map_position 11423-11423
#length     366 #checksum 7616

Entry Match      4.68; Score 137; DB 2; Length 366;
Best Local Similarity 29.0%; Pred. No. 1.93e-04;
Matches 20; Conservative 16; Mismatches 29; Indels 4; Gaps 3;

DB 157 FVYGVCCPEPHKFCLEENRPLEDQLEN--WCCRCKFCHVCGRHOATKQLIECNKR 214
OY 310 LIHSCQCSNSGHPCLDMT-MELVSMIKTYPWQCMCKCTCIICGPHHE-EMMFCDMCD 367

DB 215 NSYHPECLG 223
OY 368 RGYHTFCVG 376

RESULT 15
ENTRY      B48205      #type fragment
TITLE      All-1 protein -GRE form - mouse (fragment)
ORGANISM   Mus musculus #common_name house mouse
DATE       07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change
07-Oct-1994

ACCESSIONS B48205
REFERENCE   B48205
#authors    Ma, Q.; Alder, H.; Nelson, K.R.; Chatterjee, D.; Gu, Y.;
#journal    Nakamura, T.; Canaan, E.; Croce, C.M.; Siracusa, L.D.;
#journal    Buchberg, A.M.
#journal    Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6350-6354

```

```

#title      Analysis of the murine All-1 gene reveals conserved domains
            with human All-1 and identifies a motif shared with DNA
            methyltransferases.
#cross-references MIMD:93317679
#accession  B48205
#status     preliminary
#molecule_type mRNA
#residues   1-3866 #label HAR
#cross-references GB:L17069

GENETICS
#gene       All-1
#cross-references GB:L17069
KEYWORDS    alternative splicing; zinc finger
SUMMARY     #length 3866 #checksum 2325

Query Match      4.68; Score 137; DB 2; Length 3866;
Best Local Similarity 29.0%; Pred. No. 1.93e-04;
Matches 20; Conservative 16; Mismatches 29; Indels 4; Gaps 3;

DB 1347 FVYGVCCPEPHKFCLEENRPLEDQLEN--WCCRCKFCHVCGRHOATKQLIECNKR 1404
OY 310 LIHSCQCSNSGHPCLDMT-MELVSMIKTYPWQCMCKCTCIICGPHHE-EMMFCDMCD 367

DB 1405 NSYHPECLG 1413
OY 368 RGYHTFCVG 376

Search completed: Sun Jun 4 15:32:04 2000
Job time : 32 secs.

```

\*\*\*\*\*  
MPSRCH\_PP protein - protein database search, using Smith-Waterman algorithm  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSRCH\_PP protein - protein database search, using Smith-Waterman algorithm  
R: Sun Jun 4 15:32:58 2000: Maspar time 11.26 Seconds  
525.809 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-471-749-5  
Description: (1-410) from US09471749.pep  
Perfect Score: 3005  
Sequence: 1 MLDQVSEYLGVTSEFRKYP.....QRAPTPRVGRGKNSKEG 410

Scoring table:  
PAM 150  
Gap 11

Searched: 145341 segs, 14437480 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5LA\_COMB 2:5B\_COMB 3:6\_COMB 4:PC1\_COMB 5:backfiles1

Statistics: Mean 33.110; Variance 152.561; scale 0.217  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Score	Query	Match	Length	ID	Description	Pred. No.
1	441	14.7	405	2	US-08-881-Sequence 2, Applicatio	7.37e-28
2	147	4.9	113	4	PCR-US94-0-Sequence 37, Applicati	4.88e-03
3	147	4.9	113	1	US-08-320-Sequence 37, Applicati	4.88e-03
4	147	4.9	113	3	US-08-545-Sequence 37, Applicati	4.88e-03
5	146	4.9	187	3	US-08-545-Sequence 59, Applicati	5.82e-03
6	146	4.9	187	4	PCR-US94-0-Sequence 59, Applicati	5.82e-03
7	137	4.6	1400	4	PCR-US93-0-Sequence 7, Applicatio	2.81e-02
8	137	4.6	1400	1	US-08-080-Sequence 7, Applicatio	2.81e-02
9	130	4.3	351	4	PCR-US91-0-Sequence 1, Applicatio	9.39e-02
10	118	3.9	373	2	US-08-864-Sequence 3, Applicatio	2.69e-01
11	110	3.7	933	3	US-08-293-Sequence 2, Applicatio	2.69e-01
12	108	3.6	417	2	US-08-682-Sequence 6, Applicatio	3.72e+00
13	107	3.6	1618	1	US-07-853-Sequence 4, Applicatio	4.38e+00
14	104	3.5	218	4	PCR-US94-0-Sequence 2, Applicatio	7.11e+00
15	104	3.5	218	1	US-08-327-Sequence 2, Applicatio	7.11e+00
16	104	3.5	218	1	US-08-320-Sequence 2, Applicatio	7.11e+00
17	104	3.5	218	3	US-08-545-Sequence 2, Applicatio	6.05e+00
18	105	3.5	1376	4	PCR-US93-1-Sequence 3, Applicatio	6.05e+00
19	105	3.5	1376	4	PCR-US93-1-Sequence 3, Applicatio	6.05e+00
20	103	3.4	188	4	US-08-545-Sequence 58, Applicati	8.35e+00
21	103	3.4	188	4	PCR-US94-0-Sequence 55, Applicati	8.35e+00
22	103	3.4	1093	4	PCR-US94-0-Sequence 55, Applicati	8.35e+00
23	103	3.4	1093	3	US-08-545-Sequence 55, Applicati	8.35e+00

RESULT	ID	US-08-881-857-2	STANDARD;	PRT;	405 AA.	
24	98	3.3	781	2	US-08-675-Sequence 3, Applicatio	1.85e+01
25	98	3.3	788	2	US-08-907-Sequence 6, Applicatio	1.85e+01
26	98	3.3	997	3	US-08-872-Sequence 8, Applicatio	1.85e+01
27	96	3.2	210	2	US-09-258-Sequence 2, Applicatio	2.53e+01
28	96	3.2	210	3	US-08-569-Sequence 2, Applicatio	2.53e+01
29	96	3.2	210	2	US-09-258-Sequence 2, Applicatio	2.53e+01
30	96	3.2	250	2	US-08-867-Sequence 13, Applicati	2.96e+01
31	96	3.2	294	2	US-09-258-Sequence 10, Applicati	2.53e+01
32	96	3.2	414	1	US-07-667-Sequence 4, Applicatio	2.53e+01
33	96	3.2	593	1	US-08-296-Sequence 6, Applicatio	2.53e+01
34	96	3.2	673	2	US-08-455-Sequence 2, Applicatio	2.53e+01
35	96	3.2	704	1	US-08-188-Sequence 18, Applicati	2.53e+01
36	96	3.2	704	1	US-08-646-Sequence 18, Applicati	2.53e+01
37	96	3.2	806	1	US-07-980-Sequence 2, Applicatio	2.96e+01
38	95	3.2	861	2	US-08-343-Sequence 18, Applicati	2.96e+01
39	95	3.2	1958	1	US-07-945-Sequence 2, Applicatio	2.96e+01
40	93	3.1	104	2	US-08-710-Sequence 19, Applicati	4.03e+01
41	92	3.1	182	2	US-08-466-Sequence 90, Applicati	4.70e+01
42	92	3.1	546	2	US-09-067-Sequence 1, Applicatio	4.70e+01
43	92	3.1	651	3	US-08-650-Sequence 6, Applicatio	4.70e+01
44	92	3.1	1070	3	US-08-922-Sequence 22, Applicati	4.70e+01
45	92	3.1	1070	2	US-08-633-Sequence 2, Applicatio	4.70e+01

Sequence 1  
US-08-881-857-2  
xxxxxx

Sequence 2, Application US/08881857

Patent No. 5919660  
GENERAL INFORMATION:  
APPLICANT: KIKLY, KRISTINE R  
APPLICANT: GROSS, MITCHELL S  
APPLICANT: HURLE, MARK R  
TITLE OF INVENTION: HUMAN REQULEM  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY Forge  
STATE: PA  
COUNTRY: USA  
ZIP: 19482

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/881,857  
FILING DATE: 24-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/021,229  
FILING DATE: 26-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: ATG-50013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 405 amino acids



Query Match 4.9%; Score 147; DB 1; Length 113;  
Best Local Similarity 29.1%; Pred. No. 4,88e-03;  
Matches 23; Conservative 19; Mismatches 32; Indels 5; Gaps 4;

Dd 17 ENISKSD-FVYCCOVCEPFHFKLENERPLEDQLEN--WCCRRCFCVCGRHOAT 73  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Oy 300 ESNKKGAESLIHCQCSNSGHPSCLDWT-MEIVSMITTYMOCKECTCIIICGPHNE- 357  
::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 74 KOLECNKCNRNSYHEPCIG 92  
::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Oy 358 EMMECDCMDRGYHTFCVG 376

RESULT 4 STANDARD; PRT; 113 AA.  
ID US-08-545-86OD-37  
AC xxxxxx  
DE Sequence 37, Application US/0854586OD  
XX Sequence 37, Application US/0854586OD  
CC Patent No. 6040140  
CC GENERAL INFORMATION:  
CC APPLICANT: Croce, Carlo  
CC APPLICANT: Canaanl, Eli  
CC TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
CC TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
CC TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1 Regid  
CC NUMBER OF SEQUENCES: 94  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &  
CC ADDRESSEE: No. 6040140rIs  
CC STREET: One Liberty Place, 46th floor  
CC CITY: Philadelphia  
CC STATE: Pennsylvania  
CC COUNTRY: USA  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/545,86OD  
CC FILING DATE: 07-MAR-1996  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/04496  
CC FILING DATE: 22-APR-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US92/10930  
CC FILING DATE: 09-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/327,392  
CC FILING DATE: 19-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/320,559  
CC FILING DATE: 11-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/062,443  
CC FILING DATE: 14-MAY-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/971,094  
CC FILING DATE: 30-OCT-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/888,839  
CC FILING DATE: 27-MAY-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/805,093  
CC FILING DATE: 11-DEC-1991

```
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Deluca Esq., Mark  
CC REGISTRATION NUMBER: 33,229  
CC REFERENCE/DOCKET NUMBER: TJU-1262  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (215) 568-3100  
CC TELEFAX: (215) 568-3439  
CC INFORMATION FOR SEQ ID NO: 37:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 113 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 113 AA; 13317 MW; 64891 CN;  
  
Query Match 4.9%; Score 147; DB 3; Length 113;  
Best Local Similarity 29.1%; Pred. No. 4,88e-03;  
Matches 23; Conservative 19; Mismatches 32; Indels 5; Gaps 4;  
  
Db 17 EENISSKSD-FYVQVCCEPFHFKLENEHPLEDOLN--WCCRCKFCVCGROHOAT 73  
QY 300 ESNKKGAESLIHCOSGENSGHPSCLDMT-MELYSMIKTYPWCMECKTCIICGPAHE- 357  
Dd 74 KOLLCCNCRCNSYHPECLG 92  
QY 358 EEMMFCDMCDRGHTFCVG 376  
  
RESULT 5 STANDARD: PRT; 187 AA.  
ID US-08-545-860D-59  
AC xxxxxx  
DX Sequence 59, Application US/08545860D  
CC Patent No. 6040140  
CC GENERAL INFORMATION:  
CC APPLICANT: Croce, Carlo  
CC APPLICANT: Canaan, Eli  
CC TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
CC TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
CC TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Re  
CC NUMBER OF SEQUENCES: 94  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &  
CC ADDRESSEE: No. 604014Oris  
CC STREET: One Liberty Place, 46th floor  
CC City: Philadelphia  
CC STATE: Pennsylvania  
CC COUNTRY: USA  
CC ZIP: 19103  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/545,860D  
CC FILING DATE: 07-MAR-1996  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/04496  
CC FILING DATE: 22-APR-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US92/10930  
CC FILING DATE: 09-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/327,392  
CC FILING DATE: 19-OCT-1994  
CC PRIOR APPLICATION DATA:
```



CC APPLICATION NUMBER: US 08/320,559  
CC FILING DATE: 11-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/062,443  
CC FILING DATE: 14-MAY-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/971,094  
CC FILING DATE: 30-OCT-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/888,839  
CC FILING DATE: 27-MAY-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/805,093  
CC FILING DATE: 11-DEC-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Deluca Esq., Mark  
CC REGISTRATION NUMBER: 33, 229  
CC REFERENCE/DOCKET NUMBER: TIJ-1262  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (215) 568-3100  
CC TELEFAX: (215) 568-3439  
CC INFORMATION FOR SEQ ID NO: 59:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 187 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC HYPOTHEICAL: NO  
CC ANTI-SENSE: NO  
CC SEQUENCE 187 AA; 20805 MW; 171203 CN;

D Query Match 4.9%; Score 146; DB 3; Length 187;  
Best Local Similarity 38.0%; Pred. No. 5.82e-03;  
Matches 19; Conservative 13; Mismatches 16; Indels 2; Gaps 1;

D 8 CCIIDGECNSNVILFCDMCNLEVHOECTGVPIIPGOMLCRCLDPS 57  
I - I : : : : :  
Y 348 CIIC--GGPHHEEMAFCDMDRGVHTFCVGLAIGAPGRWICDCCORAPP 395

RESULT 6  
ID PCT-US94-04496-59 STANDARD; PRt: 187 AA.  
XX XXXXXX  
AC  
DT

Sequence 59, Application PC/TUS9404496  
GENERAL INFORMATION:  
APPLICANT: Croce, Carlo  
APPLICANT: Canani, Eli  
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &  
ADDRESSEE: Norris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496

```

CC      FILING DATE:
CC      CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Deluca Esq., Mark
CC      REGISTRATION NUMBER: 33,229
CC      REFERENCE/DOCKET NUMBER: TJU-1242
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (215) 568-3100
CC      TELEFAX: (215) 568-3439
CC      INFORMATION FOR SEQ ID NO: 59:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 187 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      HYPOTHEICAL: NO
CC      ANTI-SENSE: NO
SQ      SEQUENCE 187 AA; 20805 MW; 171203 CN;

Query Match          4.9%; Score 146; DB 4; Length 187;
Best Local Similarity 38.0%; Pred.No. 5.82e-03;
Matches 19; Conservative 13; Mismatches 16; Indels 2; Caps 1;

Db      8 CCICNDGCONSNVILFCDMCNLEHVQECGVPIIPGCHLCRCCLSPS 57
Qy      348 CIIC-GQPHEEMFCDCMCRGYHTFCVLGAIFSGRMICDCCGRAPP 395
        1 1 1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :
        1 1 1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :

RESULT   7
ID      PCT-US93-05857-7          STANDARD;          PRT;          1400 AA.
XX      AC      xxxxxx
XX      DT
XX
Sequence 7, Application PC/TUS9305857
DE
XX      Sequence 7, Application PC/TUS9305857
XX
CC      GENERAL INFORMATION:
CC      APPLICANT: Board of Regents
CC      TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
CC      TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS
CC      NUMBER OF SEQUENCES: 8
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Arnold, White & Durkee
CC      STREET: P. O. Box 4433
CC      CITY: Houston
CC      STATE: Texas
CC      COUNTRY: USA
CC      ZIP: 77210
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/US93/05857
CC      FILING DATE: 19930617
CC      CLASSIFICATION:
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: 07/900,689
CC      FILING DATE: 17/06/92
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Parker, David L.
CC      REGISTRATION NUMBER: 32,165
CC      REFERENCE/DOCKET NUMBER: AKCD:072/PAR
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (512) 320-7200
CC      TELEFAX: (512) 474-7577
CC      INFORMATION FOR SEQ ID NO: 7:
CC      SEQUENCE CHARACTERISTICS:

```

Category	Count
Conservative	20
Mismatches	16
Indels	29
Gaps	4

0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

	Query Match	4.38;	Score 130;	DB 4;	Length 351;
	Best Local Similarity	34.18;	Pred. NO. 9,39e-02;		
	Matches 15;	Conservative 10;	Mismatches 17;	Indels 2;	Gaps 1.
Db	86	CEVCGGSEIILICTCPRAHYHVCILDPFLDAPRSGKMSCPCKEK	129		









\*\*\*\*\*  
M P E R E H  
(7M)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MP\_hrp protein - protein database search, using Smith-Waterman algorithm  
Run: Sun Jun 4 15:28:16 2000; MasPar time 7.59 Seconds  
452.598 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-471-749-3  
Description: (1-238) from US09471749.pep  
Perfect Score: 1742  
Sequence: 1 MEYAKRSLSLTPKSLSRV.....GQPRKASLTPCLIKILQ 238

Scoring table: PAM 150  
Gap 11

Searched: 145341 segs, 14437480 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued

1:5A.COMB 2:5B.COMB 3:6.COMB 4:PCT.COMB 5:backfile1

Statistics: Mean 30.727; Variance 132.659; scale 0.232

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Re. No.	Score	Query Match	Length	ID	Description	Pred. No.
1	95	5.5	403	4	PCT-US95-0	Sequence 2, Applicatio 9.91e+00
2	95	5.5	403	3	US-08-989-	Sequence 4, Applicatio 9.91e+00
3	95	5.5	403	2	US-08-634-	Sequence 4, Applicatio 9.91e+00
4	95	5.5	403	2	US-08-634-	Sequence 2, Applicatio 9.91e+00
5	95	5.5	403	2	US-08-607-	Sequence 4, Applicatio 9.91e+00
6	95	5.5	403	2	US-08-607-	Sequence 2, Applicatio 9.91e+00
7	95	5.5	403	2	US-08-533-	Sequence 10, Applicatio 9.91e+00
8	95	5.5	403	2	US-08-454-	Sequence 2, Applicatio 9.91e+00
9	95	5.5	403	3	US-08-989-	Sequence 2, Applicatio 9.91e+00
10	89	5.1	431	2	US-08-665-	Sequence 5, Applicatio 2.73e+01
11	87	5.0	1015	2	US-08-680-	Sequence 32, Applicatio 3.81e+01
12	87	5.0	1130	2	US-08-519-	Sequence 6, Applicatio 4.49e+01
13	86	4.9	376	2	US-08-997-	Sequence 89, Applicatio 4.49e+01
14	86	4.9	376	2	US-08-873-	Sequence 89, Applicatio 4.49e+01
15	86	4.9	376	2	US-08-997-	Sequence 89, Applicatio 4.49e+01
16	85	4.9	497	1	US-08-633-	Sequence 6, Applicatio 5.29e+01
17	85	4.9	497	1	US-08-295-	Sequence 6, Applicatio 5.29e+01
18	85	4.9	546	2	US-08-713-	Sequence 4, Applicatio 5.29e+01
19	85	4.9	445	2	US-08-470-	Sequence 16, Applicatio 7.34e+01
20	83	4.8	170	4	PCT-US94-0	Sequence 9, Applicatio 7.34e+01
21	83	4.8	170	2	US-08-470-	Sequence 9, Applicatio 7.34e+01
22	83	4.8	170	1	US-08-081-	Sequence 8, Applicatio 7.34e+01
23	83	4.8	233	3	US-08-471-	Sequence 14, Applicatio 7.34e+01

RESULT	ID	Sequence	Score	Length	ID	Description	Pred. No.
24	83	4.8	233	1	US-08-081-	Sequence 6, Applicatio 7.34e+01	
25	83	4.8	233	4	PCT-US95-0	Sequence 24, Applicatio 7.34e+01	
26	83	4.8	233	2	PCT-US94-0	Sequence 7, Applicatio 7.34e+01	
27	83	4.8	233	2	US-08-661-	Sequence 59, Applicatio 7.34e+01	
28	83	4.8	233	1	US-08-607-	Sequence 24, Applicatio 7.34e+01	
29	83	4.8	233	1	US-08-471-	Sequence 14, Applicatio 7.34e+01	
30	83	4.8	233	2	US-08-470-	Sequence 7, Applicatio 7.34e+01	
31	83	4.8	233	1	US-08-333-	Sequence 59, Applicatio 7.34e+01	
32	83	4.8	258	1	US-07-990-	Sequence 2, Applicatio 6.23e+01	
33	84	4.8	749	3	US-08-378-	Sequence 10, Applicatio 6.23e+01	
34	84	4.8	749	3	US-08-956-	Sequence 8, Applicatio 6.23e+01	
35	84	4.8	749	4	PCT-US95-1	Sequence 8, Applicatio 6.23e+01	
36	84	4.8	749	2	US-08-852-	Sequence 8, Applicatio 6.23e+01	
37	84	4.8	749	2	US-08-820-	Sequence 8, Applicatio 6.23e+01	
38	84	4.8	749	1	US-08-369-	Sequence 8, Applicatio 6.23e+01	
39	83	4.8	1159	2	US-08-369-	Sequence 13, Applicatio 7.34e+01	
40	82	4.7	462	2	US-08-470-	Sequence 20, Applicatio 8.63e+01	
41	82	4.7	462	2	US-08-471-	Sequence 20, Applicatio 8.63e+01	
42	82	4.7	462	2	US-08-471-	Sequence 20, Applicatio 8.63e+01	
43	82	4.7	462	1	US-08-463-	Sequence 20, Applicatio 8.63e+01	
44	82	4.7	462	1	US-08-471-	Sequence 20, Applicatio 8.63e+01	
45	82	4.7	462	2	US-08-469-	Sequence 20, Applicatio 8.63e+01	

## ALIGNMENTS

RESULT	ID	Sequence	Score	Length	ID	Description	Pred. No.
1	PCT-US95-05064-2	STANDARD;	PRT;	403 AA.			
XX	XXXXXX						
XX		Sequence 2, Application PC/TUS9505064					
DE		GENERAL INFORMATION:					
CC		APPLICANT: Corixa Corporation					
CC		TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE STIMULATION AND ENHANCEME					
CC		NUMBER OF SEQUENCES: 2					
CC		CORRESPONDENCE ADDRESS:					
CC		ADDRESSEE: SEED AND BERRY					
CC		STREET: 6300 Columbia Center, 701 Fifth Avenue					
CC		CITY: Seattle					
CC		STATE: Washington					
CC		COUNTRY: USA					
CC		ZIP: 98104-7092					
CC		COMPUTER READABLE FORM:					
CC		MEDIUM TYPE: Floppy disk					
CC		COMPUTER: IBM PC compatible					
CC		OPERATING SYSTEM: PC-DOS/MS-DOS					
CC		SOFTWARE: Patent in Release #1.0, Version #1.30					
CC		CURRENT APPLICATION DATA:					
CC		APPLICATION NUMBER: PCT/US95/05064					
CC		FILING DATE: 24-APR-1995					
CC		CLASSIFICATION:					
CC		ATTORNEY/AGENT INFORMATION:					
CC		NAME: Kadlecsek, Ann T.					
CC		REGISTRATION NUMBER: P-39, 244					
CC		REFERENCE/DOCKET NUMBER: 210121.404PC					
CC		TELEPHONE: (206) 622-4900					
CC		TELEFAX: (206) 682-6031					
CC		TELEX: 3723836 SEEDANDBERRY					
CC		INFORMATION FOR SEQ ID NO: 2:					
CC		SEQUENCE CHARACTERISTICS:					
CC		LENGTH: 403 amino acids					
CC		TYPE: amino acid					
CC		TOPOLOGY: linear					
CC		MOLECULE TYPE: protein					
CC		SEQUENCE 403 AA; 45269 MW; 838361 CN;					
CC		Query Match	5.5%;	Score 95;	DB 4;	Length 403;	





CC APPLICATION NUMBER: US/08/607,509  
CC FILING DATE: 16-FEB-1996  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: MAKI, DAVID J.  
CC REGISTRATION NUMBER: 31,392  
CC REFERENCE/DOCKET NUMBER: 210121.404C3  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (206) 622-4900  
CC TELEFAX: (206) 682-6031  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 403 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 403 AA; 45326 MW; 829472 CN;

Query Match 5.5%; Score 95; DB 2; Length 403;  
Best Local Similarity 25.6%; Pred. No. 9,91e+00;  
Matches 21; Conservative 23; Mismatches 32; Indels 6; Gaps 6;

Db 230 ILVK-RESLTGEGIOFFAVEEE-HKLDLMDIVETVSIASVTPANRRKVDIAEKI 287  
OY 66 ILKRDLDLMD-KFFFLVLEDDGTVEY-EEYFALA-GDTV-FMVLQKGOKWOPSE 121  
DY 122 OGTRHPLSLSHKPAKKIDVARY 143

Db 288 NOSNHTVSMHAEKPSDRERY 309  
OY 122 OGTRHPLSLSHKPAKKIDVARY 143

RESULT 6  
ID US-08-607-509-2 STANDARD; PRT; 403 AA.  
AC xxxxxx  
DT  
DE Sequence 2, Application US/08607509  
XX Sequence 2, Application US/08607509  
CC Patent No. 5876735  
CC GENERAL INFORMATION:  
CC APPLICANT: Reed, Steven G.  
CC TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES  
CC NUMBER OF SEQUENCES: 15  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: SEED and BERRY LLP  
CC STREET: 6300 Columbia Center, 701 Fifth Avenue  
CC CITY: Seattle  
CC STATE: Washington  
CC COUNTRY: USA  
CC ZIP: 98104-7092  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent in Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/607,509  
CC FILING DATE: 16-FEB-1996  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: MAKI, DAVID J.  
CC REGISTRATION NUMBER: 31,392  
CC REFERENCE/DOCKET NUMBER: 210121.404C3  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (206) 622-4900  
CC TELEFAX: (206) 682-6031  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 403 amino acids  
CC TYPE: amino acid







CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 376 AA; 4117 MW; 720638 CN;  
Query Match 4.9%; Score 86; DB 2; Length 376;  
Best Local Similarity 22.7%; Pred. No. 4.49e+01;  
Matches 17; Conservative 20; Mismatches 34; Indels 4; Gaps 4;  
DB 14 YLAVADAFSIAPGEFFSMGLPSCGKTTTLMINGFTPEGAIKRLGADVSRTPPKR 73  
QY 82 FLVLEEDGTVEETEEYFQALA-GD-TYVMVLQKGQKQMPSEQGR-HPLSLSHKPAKI 138  
DB 74 NVNTV-FQHYALFPH 87  
QY 139 DVARTFDLYKRLNPO 153  
RESULT 14  
T US-08-873-970-89 STANDARD: PRT; 376 AA.  
AC xxxxxx  
DE Sequence 89, Application US/08873970  
XX Sequence 89, Application US/08873970  
XX Patent No. 6001361  
CC GENERAL INFORMATION:  
CC APPLICANT: Tan, Paul  
CC APPLICANT: HiYama, Jun  
CC APPLICANT: Visser, Elizabeth  
CC APPLICANT: Skinner, Margot  
CC APPLICANT: Scott, Linda  
CC APPLICANT: Prestidge, Ross  
CC TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
CC TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
CC NUMBER OF SEQUENCES: 106  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Law Offices of Ann W. Speckman  
CC STREET: 2601 Elliott Avenue, Suite 4185  
CC City: Seattle  
CC STATE: WA  
CC COUNTRY: USA  
CC ZIP: 98121  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC OPERATING SYSTEM: IBM compatible  
CC SOFTWARE: FASTSEQ for Windows Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/873,970  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/705,347  
CC FILING DATE: 29-AUG-1996  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Sleath, Janet  
CC REGISTRATION NUMBER: 37,007  
CC REFERENCE/DOCKET NUMBER: 11000.1002C1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 206-269-0565  
CC TELEFAX: 206-269-0563  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 89:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 376 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein

SQ SEQUENCE 376 AA; 4117 MW; 720638 CN;  
Query Match 4.9%; Score 86; DB 3; Length 376;  
Best Local Similarity 22.7%; Pred. No. 4.49e+01;  
Matches 17; Conservative 20; Mismatches 34; Indels 4; Gaps 4;  
DB 14 YLAVADAFSIAPGEFFSMGLPSCGKTTTLMINGFTPEGAIKRLGADVSRTPPKR 73  
QY 82 FLVLEEDGTVEETEEYFQALA-GD-TYVMVLQKGQKQMPSEQGR-HPLSLSHKPAKI 138  
DB 74 NVNTV-FQHYALFPH 87  
QY 139 DVARTFDLYKRLNPO 153  
RESULT 15  
ID US-08-997-080-89 STANDARD: PRT; 376 AA.  
AC xxxxxx  
DE Sequence 89, Application US/08997080  
XX Sequence 89, Application US/08997080  
XX Patent No. 5968524  
CC GENERAL INFORMATION:  
CC APPLICANT: WATSON, JAMES D.  
CC APPLICANT: TAN, PAUL L.J.  
CC TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICAL  
CC NUMBER OF SEQUENCES: 194  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Law Offices of Ann W. Speckman  
CC STREET: 2601 Elliott Avenue, Suite 4185  
CC City: Seattle  
CC STATE: WA  
CC COUNTRY: USA  
CC ZIP: 98121  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC OPERATING SYSTEM: IBM compatible  
CC SOFTWARE: FASTSEQ for Windows Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/997,080  
CC FILING DATE:  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Sleath, Janet  
CC REGISTRATION NUMBER: 37,007  
CC REFERENCE/DOCKET NUMBER: 11000.1007  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 206-269-0565  
CC TELEFAX: 206-269-0563  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 89:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 376 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 376 AA; 4117 MW; 720638 CN;  
Query Match 4.9%; Score 86; DB 2; Length 376;  
Best Local Similarity 22.7%; Pred. No. 4.49e+01;  
Matches 17; Conservative 20; Mismatches 34; Indels 4; Gaps 4;  
DB 14 YLAVADAFSIAPGEFFSMGLPSCGKTTTLMINGFTPEGAIKRLGADVSRTPPKR 73

Sun Jun 4 17:48:41 2000

US-09-471-749-3.rtf

Page 8

QY 82 FLVLEBGTIVETETETQALA-GD-TVFMVLQKQKQKQPPSEQTR-NPLSLHKRAKKI 138

DB 74 NVNTV-FQHYALPPH 87

QY 139 DVARVTFDLXKLNQ 153

Search completed: Sun Jun 4 15:28:28 2000  
Job time : 12 secs.

45 92 1070 2 US-08-633- Sequence 2, Applicatio 4.70e+01

ALIGNMENTS

RESULT 1 US-08-881-857-2 STANDARD: PRT: 405 AA.

Sequence 2, Application US/08881857  
Patent No. 5919660  
GENERAL INFORMATION:  
APPLICANT: KIRBY, KRISTINE K  
APPLICANT: GROSS, MITCHELL S  
APPLICANT: HURLE, MARK R  
TITLE OF INVENTION: HUMAN REQUIM  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RAINIER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/881,857  
FILING DATE: 24-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/021,229  
FILING DATE: 26-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: ATG-50013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 405 amino acids

Sun Jun 4 17:48:43 2000

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 405 AA: 45780 MW: 803227 CN:  
14.7% Score 441: DB 2; Length 405;  
Best Local Similarity 41.0%; Pred. No. 7.37e-28;  
Matches 57: Conservative 30: Mismatches 48: Indels 4: Gaps 3:

DB 250 EGEDEKEDSOPPTPVSRSEKSKSGKGGDGLALPNNYCDFCLGDSKINKKKTGQPEELVSC 309  
256 QRSKDKAAT-PRKDGPKRSVLSKSVFPGYKPKVAPNAICIGICL-KGKSNKKKGKAKAESLIHC 313  
DB 310 SDGSRSGHPSCLQFTPVMAAAVKTYYRMOCKCNCNICGTSENDDOLLFCDDCDRGYHMY 369  
314 SOCENSGHPSCLDMTMTLVSIMIKTYIPWOCMECKTCIICGQPHHEEEMFCCDMCDRGYHMY 373  
DB 370 CLTPSMSPPEEGSWSCNHC 388  
374 CV--GIGAIPIPSGRWICDCC 390

RESULT 2 US-08-881-857-2 STANDARD: PRT: 405 AA.

\*\*\*\*\*

```

Mpsrch_pp  protein - protein database search, using Smith-Waterman algorithm
R1 1: Sun Jun 4 15:27:04 2000; MspPar time 17.96 Seconds
Tabular output not generated. 625.036 Million cell updates/sec

```

Scoring table: PAM 150

Post-processing: Minimum Match 08 ~  
Listing first 45 summaries

Database: - plr62

Statistics: Mean 45.595; Variance 83.289; scale 0.547

**Pred.** No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Rd	Query	Score	Match Length	ID	Description	Pred. No
----	-------	-------	--------------	----	-------------	----------

1	1390	79.8	239	2	A42445	fat-specific protein	9.98e-28
2	102	5.9	1405	2	T04426	hypothetical protein	1.80e-01
3	101	5.8	335	2	G70107	hypothetical protein	2.48e-01
4	100	5.7	692	2	H70362	glycogen phosphorylase	3.40e-01
5	98	5.6	435	2	H70929	hypothetical protein	6.35e-01
6	97	5.6	919	2	B17265	hypothetical protein	8.65e-01
7	95	5.5	200	2	F70900	hypothetical protein	1.59e+00
8	95	5.5	403	2	T02861	EIF-4A protein - Iais	1.59e+00
9	95	5.5	413	2	S38358	translational initiatio	1.59e+00
10	95	5.5	1252	2	T14272	translational initiatio	1.59e+00
11	94	5.4	341	2	C12403	coactin-binding pro	1.59e+00
12	94	5.3	214	2	T08922	hypothetical protein	2.16e+00
13	92	5.3	254	2	S52021	probable calcineurin	3.92e+00
14	92	5.3	413	2	S52018	translational initiatio	3.92e+00
15	92	5.3	414	2	JT08839	translational initiatio	3.92e+00
16	92	5.3	506	2	F89867	two-component sensor	3.92e+00
17	92	5.3	562	2	A55281	carboxylesterase (EC	3.92e+00
18	93	5.3	633	2	S44795	P90G8.3 protein - Cae	3.91e+00
19	92	5.3	1049	1	T19454	ATP-dependent permeas	3.92e+00
20	92	5.3	1179	2	S14584	AMP resistance proteol	3.92e+00
21	91	5.2	251	2	A69020	ABC transporter subun	5.26e+00
22	90	5.2	261	2	T00319	hypothetical 27k prot	7.10e+00
23	91	5.2	298	2	T01948	hypothetical protein	5.26e+00

45	87	5.0	3951	1	VFHBI	hypothetical protein	7.04e+000
44	88	5.1	348	2	CG4576	hypothetical protein	5.26e+000
43	87	5.0	1468	2	CG4576	hypothetical protein	5.26e+000
42	88	5.1	1179	2	CG4576	hypothetical protein	5.26e+000
41	88	5.1	1174	2	CG4576	hypothetical protein	5.26e+000
40	89	5.1	1102	2	CG4576	hypothetical protein	5.26e+000
39	89	5.1	1082	2	CG4576	hypothetical protein	5.26e+000
38	89	5.1	896	2	CG4576	hypothetical protein	5.26e+000
37	89	5.1	896	2	CG4576	hypothetical protein	5.26e+000
36	88	5.1	870	2	CG4576	hypothetical protein	5.26e+000
35	88	5.1	858	2	CG4576	hypothetical protein	5.26e+000
34	89	5.1	587	2	CG4576	hypothetical protein	5.26e+000
33	89	5.1	469	2	CG4576	hypothetical protein	5.26e+000
32	88	5.1	412	2	CG4576	hypothetical protein	5.26e+000
31	88	5.1	405	2	CG4576	hypothetical protein	5.26e+000
29	89	5.1	301	2	CG4576	hypothetical protein	5.26e+000
28	88	5.1	261	2	CG4576	hypothetical protein	5.26e+000
27	90	5.2	2285	2	CG4576	hypothetical protein	5.26e+000
26	91	5.2	561	2	CG4576	hypothetical protein	5.26e+000
25	90	5.2	316	2	CG4576	hypothetical protein	5.26e+000
24	90	5.2	301	2	CG4576	hypothetical protein	5.26e+000
23	91	5.2	316	2	CG4576	hypothetical protein	5.26e+000
22	91	5.2	674	2	CG4576	hypothetical protein	5.26e+000
21	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
20	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
19	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
18	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
17	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
16	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
15	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
14	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
13	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
12	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
11	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
10	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
9	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
8	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
7	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
6	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
5	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
4	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
3	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
2	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000
1	90	5.2	674	2	CG4576	hypothetical protein	5.26e+000

## RESULT 1

```
ENTRY      A42445      #type complete
TITLE      fat-specific protein_FSP27 - mouse
ORGANISM   Mus musculus #common_name house mouse
DATE       04-Mar-199399 #sequence_revision 18-Nov-1994 #text_change
           11-Apr-1995
```

#journal	#authors	REFERENCE	ACCESSIONS
		A42445	A42445
		A42445	A42445
		Daneesh, U.; Hoesch, W.; Ringold, G.M	
		J Biol Chem (1992) 267:185-193	
#title		Cloning and transcriptional regulation	

**#title**  
Cloning and transcriptional regulation of a novel  
adipocyte-specific gene, *EBP27*. C/EBP-  
protein (C/EBP) and C/EBP-like proteins interact with  
sequences required for differentiation-dependent  
expression.

```
##cross-references MIMD:92202389
#accession A42445
##status preliminary
##molecule_type mRNA
##residues 1-239 ##label DAN
##experimental_source TAI cells
##note sequence extracted from NCBI database/MPBRN-923164
```

```
#length 239 #molecular-weight 27292 #checksum 3042
```

Query Match	79.88;	Score 1390;	DB 2;	Length 239;
Best Local Similarity	79.18;	Pred. No. 9	98e-281;	

Db	1	MDYAMKSLILTYRSLRSRHAASVAVYVQOLYSKRSRETPARCRSTADRYKRGIMA	60
		:     :     :     :     :     :     :     :     :     :	
Qy	1	MEYAMKSLILTYKSLSRHVSRVTSVYVQOLLSESPAPAPARCSTADRSYRGIMA	60
Db	61	HSLEDLLINKVODILKDKDPESLVLEEDGTVETEERYFOALAKDYMVILKQOKRPPS	120
		:     :     :     :     :     :     :     :     :	
Qy	61	YSLLEDLLIKRDPILMLADRPFLVLEEGETVETEERYFOALAGDTPVMVLQKQKOMPSS	120
Db	121	EQRKRAQALILSQPKRPKIDVARVTFEDLYKLNPOBETICLNKRYKILBYTSLSYDLHCPR	180
		:     :     :     :     :     :     :     :     :	
Qy	121	EQRGRHP-LSLSRKPAPKIDVARVTFEDLYKLNPOBETICGLNRYKTFIDYLSYDLHCQG	178
Db	181	AKRIEYIARVILFSSQALGHHLLIGSSYMOOFLDATTEDQPAARAKSSLLPACILMLO	239
		:     :     :     :     :     :     :     :     :	
Qy	180	AKRIEAFARWALFSSQALGHHLLIGSSYMOOFLDATTEDQGPAPGRKASSLLPACILKILQ	238



ENTRY T04426 #type complete  
TITLE hypothetical protein T18B16.20 - Arabidopsis thaliana  
ALTERNATE\_NAMES hypothetical protein F13C5.220  
ORGANISM #formal\_name Arabidopsis thaliana #common\_name mouse-ear  
cress  
DATE 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change  
ACCESSIONS T04426; T05042  
REFERENCE 215359  
#authors Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.  
#submission #submitted to the Protein Sequence Database, April 1998  
#accession T04426  
#molecule\_type DNA  
#residues 1-1405 #label BEV  
#cross-references EMBL:AL021687  
#experimental\_source cultivar Columbia; BAC clone T18B16  
REFERENCE 215395  
#authors Bevan, M.; Pohl, T.; Welzenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.  
#submission submitted to the Protein Sequence Database, February 1998  
#accession T05042  
#molecule\_type DNA  
#residues 1-1405 #label BEW  
#cross-references EMBL:AL021711  
#experimental\_source cultivar Columbia; BAC clone F13C5  
GENETICS  
#map\_position 4  
#introns 893/3; 1164/2; 1183/3; 1192/3; 1198/2; 1214/3; 1251/1;  
1282/2; 1327/1; 1381/3  
T18B16.20; F13C5.220  
SUMMARY #length 1405 #molecular-weight 160485 #checksum 8041  
#note 5.98; Score 102; DB 2; Length 1405;  
Query Match Best Local Similarity 27.3%; Pred. No. 1.80e-01;  
Matches 12; Conservative 15; Mismatches 16; Indels 1; Gaps 1;  
DB 95 LQYPLESLKEKIKKRNKLLILIDEGSMTEEDVMQEL 138  
QY 59 MAFSLDILLKVDLMADKP-FLVLLEDGTVEETEEYFQAL 101  
RESULT 3  
ENTRY G70107 #type complete  
TITLE hypothetical protein BB0063 - Lyme disease spirochete  
ORGANISM #formal\_name Borrelia burgdorferi #common\_name Lyme disease  
spirochete  
DATE 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change  
ACCESSIONS G70107  
REFERENCE A70100  
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uitterlinden, T.; Watney, L.; McDonald, L.; Artach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.  
#journal Nature (1997) 390:580-586  
#title Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
#accession G70107  
#cross-references M01D:98065943  
#status preliminary; nucleic acid sequence not shown;  
#molecule\_type DNA  
#residues 1-335 #label KLE  
#cross-references GB:AE001119; GB:AE000783; NID:92687936; PID:92687947;  
TIGR:BB0063  
SUMMARY #experimental\_source strain B31  
#length 335 #molecular-weight 37305 #checksum 7981

Query Match 5.8%; Score 101; DB 2; Length 335;  
Best Local Similarity 33.3%; Pred. No. 2.48e-01;  
Matches 17; Conservative 15; Mismatches 17; Indels 2; Gaps 2;  
DB 26 TIKALLVIFGSLVFAFVLENNETVVPNYSLAE-DAV-LELOR 74  
QY 62 SLEDLLKVDLMADKPFLVLEEDGTVEETEEYFQALAGDYVMVQK 112  
RESULT 4  
ENTRY H70362 #type complete  
TITLE glycogen phosphorylase - Aquifex aeolicus  
ORGANISM #formal\_name Aquifex aeolicus  
DATE 08-May-1998 #sequence\_revision 08-May-1998 #text\_change  
ACCESSIONS H70362  
REFERENCE A70300  
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Anjaj, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.  
#journal Nature (1998) 392:353-358  
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
#cross-references M01D:98196666  
#accession H70362  
#status preliminary; nucleic acid sequence not shown;  
#introns translation not shown  
#molecule\_type DNA  
#residues 1-692 #label AOF  
#cross-references GB:AE000704; NID:92983301; PID:92983305; GB:AE000657  
#experimental\_source strain VF5  
GENETICS  
#gene glpP  
SUMMARY #length 692 #molecular-weight 81158 #checksum 7682  
Query Match 5.7%; Score 100; DB 2; Length 692;  
Best Local Similarity 21.7%; Pred. No. 3.40e-01;  
Matches 18; Conservative 25; Mismatches 36; Indels 4; Gaps 4;  
DB 230 LRLROQIVAGFVILVLEKIDGAFHINEDYSPVFLAEIFKLKGLTMDKALEYR 289  
QY 68 LKVDRLMLA-DKPFPLV-LEEDGTVEETEEYFQALAG-DYFVVLQKQKMPSEQGT 124  
DB 290 KISLFTTHPLR-VAVNTYFPM 311  
QY 125 RHPLSLSHKPAKKIDVAVYFPL 147  
RESULT 5  
ENTRY A70929 #type complete  
TITLE hypothetical protein RV1783 - Mycobacterium tuberculosis  
ORGANISM #formal\_name Mycobacterium tuberculosis  
DATE 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change  
ACCESSIONS A70929  
REFERENCE A70500  
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekle, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Hamlin, N.; Holroyd, S.; Hornsby, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Ratandream, M.A.; Rogers, S.; Squires, R.; Sulston, J.E.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
#journal Nature (1998) 393:537-544  
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
#cross-references M01D:98295987

[illegible]

```

Feltwell, T.; Gentile, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quall, M.A.;
Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skilton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
Nature (1996) 393:537-544
Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.

#journal
#title
#cross-references MUID:98295987
#accession F70900
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule-type DNA
##residues 1-200 #label COL
##cross-references GB:280108; PID:AL123456; NID:g3256012; PID:e1301423;
PID:g3256014
GENETICS
##experimental-source strain H37Rv
CLASSIFICATION
#gene Rv1401
#superfamily Mycobacterium tuberculosis hypothetical protein
Rv1401
SUMMARY
#length 200 #molecular-weight 21257 #checksum 7966
Query Match 5.5%; Score 95; DB 2; Length 200;
Best Local Similarity 26.7%; Pred. No. 1.59e+00;
Matches 12; Conservative 16; Mismatches 15; Indels 2; Gaps 2;

Db 51 WA-FYF-GIGAFILHICFIALPLPARAPBGRGVAAYVAMCV 93
190 WAFSMAQARGHVLGTSTCYLQQLDNTGEGQPKRASSLIPCL 234
OY

RESULT 8
ENTRY T02861 #type complete
TITLE ERF-4A protein - Leishmania major
ORGANISM #formal_name Leishmania major
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
26-Aug-1999
ACCESSIONS T02861; T02862
REFERENCE 214740
#authors Myler, P.J.
#submission submitted to the EMBL Data Library, May 1998
#description The nucleotide sequence of Leishmania major Friedlin
chromosome 1.
#accession T02861
##status preliminary
##molecule-type nucleic acid
##residues 1-403 #label MYL
##cross-references EMBL:AE001274; NID:g3264850; PIDN:AAC24684.1;
PID:g3002484
#accession T02862
##status preliminary
##molecule-type nucleic acid
##residues 1-403 #label MY2
##cross-references EMBL:AE001274; NID:g3264850; PIDN:AAC24685.1;
PID:g3002483
CLASSIFICATION
#superfamily translation initiation factor eIF-4a; DEAD/H box
helicase homology
KEYWORDS ATP; P-loop
FEATURE
70-360 #domain DEAD/H box helicase homology #label DEAD\
70-77 #region nucleotide-binding motif A (P-loop)\
174-179 #region nucleotide-binding motif B\
178-181 #region DEAD motif
SUMMARY #length 403 #molecular-weight 45326 #checksum 2708
Query Match 5.5%; Score 95; DB 2; Length 403;
Best Local Similarity 25.6%; Pred. No. 1.59e+00;
Matches 21; Conservative 23; Mismatches 32; Indels 6; Gaps 6;

Db 230 ILVY-RESLTLEGIRQFPAVEEE-HKLDPTLMDLVTSTVIAQSVIFANTRRRKYDWAIEKL 287

```

[illegible]

```

#cross-references EMBL:AL078465
GENETICS
#experimental_source cultivar Columbia; BAC clone T15N24
#map_position 4
#introns 41/3; 57/3; 77/3; 114/1; 158/3; 196/2
#note T15N24.10
SUMMARY
#length 214 #molecular-weight 24404 #checksum 7497

Query Match
#score 5.3%; Score 92; DB 2; Length 214;
Best Local Similarity 29.7%; Pred. No. 3.92e+00;
Matches 19; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

Db
130 OGVR 133
159 LNVK 162

RESULT 13
ENTRY S52021 #type fragment
TITLE translation initiation factor eIF-4A.6 - common tobacco
#fragment
ORGANISM #formal_name Nicotiana tabacum #common_name common tobacco
DATE 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change
26-Aug-1999
ACCESSIONS S52021; S44932
REFERENCE S52017
#authors Owttrim, G.W.; Mandel, T.; Trachsel, H.; Thomas, A.A.M.;
#journal Kuhlmeier, C. (1994) 26:1747-1757
#title Plant Mol. Biol. (1994) 26:1747-1757
#cross-references MUID:95161700
#accession S52021
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-254 #label OMT
#cross-references EMBL:X79139
REFERENCE S44928
#authors Owttrim, G.W.; Mandel, T.; Trachsel, H.; Adrl, M.A.;
#journal Kuhlmeier, T.; Kuhlmeier, C.
#title submitted to the EMBL Data Library, May 1994
#cross-references MUID:95161700
#accession S44932
#molecule_type mRNA
#residues 1-58, 'I', 60-253 #label OMT
#cross-references EMBL:X79139; NID:9485946; PIDN:CAA55740.1;
CLASSIFICATION #superfamily translation initiation factor eIF-4A; DEAD/H box
#molecule_type mRNA
#residues 1-254 #label OMT
#cross-references EMBL:X79139; NID:9485946; PIDN:CAA55740.1;
KEYWORDS #superfamily translation initiation factor eIF-4A; DEAD/H box
FEATURE #helix
1-216 #domain DEAD/H box helicase homology (fragment) #label
26-31 #region nucleotide-binding motif B\
30-33 #region DEAD motif
SUMMARY #length 254 #checksum 2492

Query Match
#score 5.3%; Score 92; DB 2; Length 254;
Best Local Similarity 26.8%; Pred. No. 3.92e+00;
Matches 19; Conservative 20; Mismatches 27; Indels 5; Gaps 5;

Db
82 ILVK-RDELTTLEGIQFVYNDKEMKLETTCDLYETLAITQSVTFVNTRRKVDWLTDRM 140
66 ILTKYRDITMLAD-KPFLVLEEDGTVETE-EYFOALA-GDTV-FMVLQKQKQMPSE 121
141 RGRDHTVSATH 151
122 QGTRHPLSLSH 132

```

```

RESULT 14
ENTRY S52018 #type complete
TITLE translation initiation factor eIF-4A.11 - common tobacco
ORGANISM #formal_name Nicotiana tabacum #common_name common tobacco
DATE 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change
26-Aug-1999
ACCESSIONS S52018; S44928
REFERENCE S52017
#authors Owttrim, G.W.; Mandel, T.; Trachsel, H.; Thomas, A.A.M.;
#journal Kuhlmeier, C.
#title Plant Mol. Biol. (1994) 26:1747-1757
#cross-references MUID:95161700
#accession S52018
#molecule_type mRNA
#residues 1-413 #label OMT
#cross-references EMBL:X79136
#note only a part of the nucleic acid sequence is shown
REFERENCE S44928
#authors Owttrim, G.W.; Mandel, T.; Trachsel, H.; Adrl, M.A.;
#journal Kuhlmeier, T.; Kuhlmeier, C.
#title submitted to the EMBL Data Library, May 1994
#cross-references MUID:95161700
#accession S44928
#molecule_type mRNA
#residues 1-39, 'E', 41-97, 'I', 99-413 #label OMT
#cross-references EMBL:X79136; NID:9485986; PIDN:CAA55737.1;
CLASSIFICATION #superfamily translation initiation factor eIF-4A; DEAD/H box
#molecule_type mRNA
#residues 1-39, 'E', 41-97, 'I', 99-413 #label OMT
#cross-references EMBL:X79136; NID:9485986; PIDN:CAA55737.1;
KEYWORDS #superfamily translation initiation factor eIF-4A; DEAD/H box
FEATURE #helix
84-372 #domain DEAD/H box helicase homology #label DEAD\
84-91 #region nucleotide-binding motif A (P-loop)\
185-190 #region nucleotide-binding motif B\
189-192 #region DEAD motif
SUMMARY #length 413 #molecular-weight 46873 #checksum 2493

Query Match
#score 5.3%; Score 92; DB 2; Length 413;
Best Local Similarity 26.8%; Pred. No. 3.92e+00;
Matches 19; Conservative 20; Mismatches 27; Indels 5; Gaps 5;

Db
241 ILVK-RDELTTLEGIQFVYNDKEMKLETTCDLYETLAITQSVTFVNTRRKVDWLTDRM 299
66 ILTKYRDITMLAD-KPFLVLEEDGTVETE-EYFOALA-GDTV-FMVLQKQKQMPSE 121
300 RGRDHTVSATH 310
122 QGTRHPLSLSH 132

RESULT 15
ENTRY JN0839 #type complete
TITLE translation initiation factor eIF-4A - wheat
ORGANISM #formal_name Triticum aestivum #common_name wheat
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change
22-May-1998
ACCESSIONS JN0839
REFERENCE JN0839
#authors Metz, A.M.; Browning, K.S.
#journal Gene (1993) 131:299-300
#title Sequence of a cDNA encoding wheat eukaryotic protein
#cross-references MUID:94010325
#accession JN0839
#molecule_type mRNA
#residues 1-414 #label MET
COMMENT This protein participates in the ATP-dependent unwinding of mRNA in
the presence of eukaryotic translation factor 4B and 4F prior to
the binding of mRNA to the 40S ribosomal subunits.
CLASSIFICATION #superfamily translation initiation factor eIF-4A; DEAD/H box
#molecule_type mRNA
#residues 1-414 #label MET
KEYWORDS #superfamily translation initiation factor eIF-4A; DEAD/H box
ATP; P-loop; protein biosynthesis; RNA binding

```

FEATURE  
85-373 #domain DEAD/H box helicase homology #label DEAD\  
85-92 #region nucleotide-binding motif A (P-loop)\  
186-191 #region nucleotide-binding motif B\  
190-193 #region DEAD motif\  
91 #binding\_site ATP (Lys) #status predicted  
SUMMARY #length 414 #molecular-weight 46927 #checksum 2762

Query Match 5.3%; Score 92; DB 2; Length 414;  
Best Local Similarity 26.8%; Pred. No. 3.92e+00;  
Matches 19; Conservative 20; Mismatches 27; Indels 5; Gaps 5;

Db 242 ILVK-RDELTLEIGIKQFYVNEKEKIDTLCIDLYETLAITQSIFVNTRRKVDWLTDM 300  
QY 66 LILKVDITLMD-KPFLVLEEDGTVEE-EYQALA-GDIV-FMYLQKQKQPPSE 121  
Db 301 RGRDHTVSATN 311  
QY 122 QGTRHPLSLSH 132

Se completed: Sun Jun 4 15:27:26 2000  
Job time : 22 secs.